What is in the dataset? [10 April 2017]

Protein structures as generated by the flexibility analysis

The dataset contains subfolders, named after the PDB codes of the proteins such as “ERp57” or classes such as “hPDI” (human PDI).

Typically, these folders contain a few files related to running FIRST and/or normal mode analysis. A file called [PDBcode]_FH.ps will be a rigidity dilution plot. In addition, two subfolders: “Runs” and “Modes”. The “Modes” folder just contains the normal mode vectors for modes 07 – 11. The “Runs” folder contains a folder called “2.0” (all simulations included here have been run at Ecut = - 2.0 kcal/mol.)

The Runs/2.0/ folder contains a set of folders called e.g. Mode07-pos or Mode11-neg. The contents of these are the pdb files saved from the FRODA runs in the mode and direction given in the folder name.

Other files are standard files produced form running FIRST/FRODA and possibly some extra ones we’ve generated in the process of analyzing the FRODA runs.

Some of the folders contain additional subfolders when the initial PDB structure contains two protein molecules in the crystal unit, and we have a folder called e.g. 4EKZ_chainA, which itself will contain the Runs/2.0/Mode directories.

Movies

The files named *-movies.zip contain movies made from the conformers in the other folders. After unzipping, one gets a complete set of such movies, made with Pymol, for Ecut= -1.0 and -2.0, modes 7 to 11 and in positive and negative direction. Also, all movies have been made with “cartoon” and “sphere” setting in pymol.

Additional information

Here is a list of some of the key residues used in generating many of the graphs.

- Central atoms to represent the domain:

2B5E: Ala86, Val191, Val291, Ala433 (CA atom numbers 981, 2583, 4147, 6382).

3BOA: same as 2B5E (CA atom numbers 946, 2547, 4111, 6344).

3F8U: Lys82, Ala185, Ala300, Ala432. (CA atom numbers 858, 2420, 4279, 6369)

4EKZ: Ala80, Gly185, Ile289, Ala 423. (CA atom numbers 1175, 2739, 4331, 6508)

4EL1: Same as 4EKZ (CA atom numbers 961, 2525, 4127, 6270) (=4EL1chainA)

- Active site residues:

yPDI: Cys61 and Cys406 (first C residue in CGHC bit). (b and bp "active sites" are 170 and 266)

hPDI: Cys53 and Cys397. (b and bp "active sites" are D167 and D267)
(4EL1_chainA atom numbers of these are 564, 2256, 3778, 5855)
(4EKZ atom numbers of these are 776, 2470, 3982, 6093)

ERp57: Cys57 and Cys406. (b and bp "active sites" are E166 and A273)
(ERp57_chain A atom numbers of these are 408, 2112, 3828, 5952).

- Four atoms per domain (central to the beta-sheets, like in the yPDI paper)

Residue numbers are given (alpha-carbon atoms were used in calculations).

yPDI:
a: 53, 55, 108, 110.
b: 163, 165, 202, 204.
bp: 260, 262, 314, 316.
ap: 398, 400, 453, 455.
yPDI fragment (2K18):
b: 28, 30, 74, 76.
bp: 128, 130, 186, 188.

hPDI AND ERp57:
a: 49, 51, 103, 105.
b: 156, 158, 202, 204.
bp: 261, 263, 326, 328.
ap: 398, 400, 453, 455.

ERp44:
a: 21, 23, 81, 83.
b: 134, 136, 179, 181.
bp: 235, 237, 292, 294.

ERp27:
b: 59, 61, 106, 108.
bp: 164, 166, 222, 224.

- Missing residues

In hPDI there are some missing residues in the PDB files. This is true for both structures, although different residues are missing in each case.

hPDI reduced full length. PDB code 4EKZ

1 chain.
Missing residues: 322,323, 240-244 (inclusive)
Residue 321 and 324 have CA atoms 4861 and 4877. MAX dist in all ten mode-directions is 9.92116
Residue 239 and 245 have CA atoms 3611 and 3627. MAX dist in all ten mode-directions is 19.8494

hPDI oxidized fill length. PDB code 4EL1.

2 chains, A and B
Missing residues: 320-323, 250-254.
Residue 249 and 255 have CA atoms 3557 and 3578. MAX dist in all ten mode-directions is 10.2586
Residue 319 and 324 have CA atoms 4624 and 4639. MAX dist in all ten mode-directions is 6.509444

From part of a protein with 2 CA atoms in a line (an N terminus), the separation of CAs there is 3.87.
Between atom 4681 and 4877 there are 3 such separations => max possible is 3*3.87 = 11.61.
Between atom 3611 and 3627 there are 6 such separations => max possible is 6*3.87 = 23.22.
Between atom 3557 and 3578 there are 6 such separations => max possible is 6*3.87 = 23.22.
Between atom 4626 and 4639 there are 5 such separations => max possible is 5*3.87 = 19.35.